14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

COVARIANCE-BASED FREQUENCY REASSIGNMENT FOR DETECTION OF NON-STATIONARY SINUSOIDS

Miroslav Zivanovic, Javier Francés

Dept. Electrical Engineering, Universidad Pública de Navarra, Campus Arrosadía, 31006 Pamplona, Spain
phone: +34 948 169024, fax: +34 948 169720, email: miro@unavarra.es
web : www.unavarra.es

ABSTRACT

A new approach to the frequency reassignment for sinusoidal energy localization in the time-frequency plane is presented. While the classical reassignment method operates statically on each point in the plane, a more global approach focuses on the dynamics of the frequency reassignment operator along the frequency grid. It is shown that the first mixed moment of the frequency reassignment operator and linear frequency well describes the Discrete-Time Fourier Transform (DTFT) spectral peaks in terms of sinusoids and noise. The new method eludes spurious artifacts in noisy regions, produced by the classical reassignment method randomly clustering noise. The proposed method is shown to have very good behavior for low SNR and it is computationally efficient.

1. INTRODUCTION

The detection of time-varying sinusoids buried in noise is a milestone for musical signal analysis/synthesis. The basis for all the methods is always some time-frequency representation (TFR) among which the spectrogram, defined as a squared modulus of the Short Time Fourier Transform (STFT) is the most prominent. It represents the signal's energy density obtained by averaging the energy delimited by a time-frequency analysis window and assigning that value to the geometric center of the window. As, in general, the energy spectral density is not uniform, the spectrogram often provides for a poor localization of time-frequency components of the analyzed signal.

In order to improve the readability of the spectrogram, the reassignment method was proposed [1]. It modifies the spectrogram by shifting the energy assigned to the geometric center of the window to the center of gravity of that energy. The new time-frequency coordinates are the group delay and instantaneous frequency respectively and the shift is determined by the time and frequency reassignment operators (t_r,ω_r). The reassigned spectrogram possesses the properties of time and frequency invariance and energy conservation and it perfectly localizes pure tones, chirps and impulses. In spite of that, the reassignment method has some inherent shortcomings. On one hand, the presence of broad-band noise provides for a non-zero reassignment vector in the regions where there are no genuine components. A method of supervised reassignment [2] analyzes the statistics of the reassignment vector [3] through multi-window STFTs in order to determine the presence of true sinusoids. It seems to work well for isolated chirp signals but it is computationally quite expensive. On the other hand, the frequency reassignment operator has been used as a sinusoidal estimator [4]. Although it was proven to be very good in terms of the Cramer-Rao bound, the bias removal for multiple tones shows to be a very difficult task [5].

In this paper we present a new approach to sinusoidal energy localization by means of the covariance between the frequency reassignment operator and linear frequency. Unlike the aforementioned methods, neither the reassignment spectrogram is calculated nor is the frequency reassignment operator used as a sinusoidal estimator. We rather investigate into the variation of the frequency reassignment operator for a set of points clustered within a DTFT modulus peak. This approach, being an extension to the reassignment method, brings out more information about the signal's energy spectral density. The sinusoidal and noise peaks are shown to have respective probability density functions well separated along the covariance grid, the fact which relaxes the threshold determination conditions. Once the statistics are generated, they can be used for sinusoidal detection in musical signals. The main advantages of the proposed method are absence of spurious artifacts in purely noise regions, robustness to low SNR and high computational efficiency.

The reminder of this paper is organized as follows. In Section 2 the covariance method is presented and evaluated for sinusoidal and noise DTFT peaks. Section 3 is dedicated to the statistical characterization of sinusoids and noise through the covariance method. In Section 4 we present some experimental results. Finally, some concluding remarks are given in Section 5.

2. DESCRIPTION OF THE COVARIANCE METHOD

For a DTFT spectral peak of width $\Delta\omega$ (distance between two contiguous local minima), the covariance of the frequency reassignment operator ω_r and linear frequency ω is given as:

$$C_{\omega\omega_r} = \langle \omega\omega_r \rangle - \langle \omega \rangle \langle \omega_r \rangle \ , \qquad (1)$$

where all the averages are calculated with respect to the energy spectral density function $P(\omega)$ of the windowed signal $x(n)$:

$$P(\omega) = |X(\omega)|^2 \Big/ \int_{\Delta\omega} |X(\omega)|^2 \ , \quad X(\omega) = DTFT\{x(n)\} . \qquad (2)$$

The covariance $C_{\omega\omega r}$ is the first mixed moment and it reflects in a simple way the strength of the dependence. Also, it gives a gross indication of where the energy density is concentrated with respect to the product $\omega\omega_r$.

The reassignment time-frequency operator (t_r, ω_r) can be interpreted as a gradient to the STFT, following the steepest-descent direction. That is to say, the reassignment operator shows the direction of maxima in the STFT modulus. It was shown that for the Gaussian window of unit variance this is exactly true while for other analysis windows there is an additional "non-analyticity" factor [6]. Therefore, ω_r can be conceived as a frequency component of the STFT gradient vector. By making use of [1] we give an alternative definition for ω_r:

$$\omega_r = \frac{N}{2\pi} \frac{|X_{dt}(\omega)|}{|X(\omega)| + a} \sin[\arg X_{dt}(\omega) - \arg X(\omega)] , \qquad (3)$$

where $X_{dt}(\omega)$ is the DTFT of $x(n)$ multiplied by the time derivative of the analysis window and N is the number of FFT bins. In expression (3) the modulus can be roughly viewed as the gradient intensity and the sine as the gradient direction. The parameter a, which modifies the original expression from [1], makes (3) more robust against ambiguous results which often arise in the neighborhood of local minima in $|X(\omega)|$. For a spectral peak in general, while ω is linear, we expect ω_r to be zero at the peak maximum and to increase in both directions in some fashion up to the moment when it starts decreasing towards zero in the local minima. In the following subsections we analyze this behavior for sinusoidal and noise peaks and establish the relation among $\Delta\omega$, ω_r and $C_{\omega\omega r}$.

2.1. Sinusoids

For stationary sinusoids the reassignment shift is completely determined by ω_r because the center of gravity always coincides with the geometric center of the analysis window. Therefore, ω_r is a linear function of frequency around the peak maximum (instantaneous frequency) with the slope equal to 1. For non-stationary sinusoids $t_r \neq 0$, i.e. both components of the reassignment shift will contribute to the gradient. For linear chirps ω_r will still vary in a linear fashion, but the slope will depend on the modulation strength: the stronger the modulation the less pronounced is the slope. This occurs because the center of gravity is shifted in both time and frequency from the geometric center of the window. For amplitude-modulated sinusoids both components of the gradient will be non zero but as the instantaneous frequency is constant, ω_r will be the same as for stationary sinusoids. In presence of mixed AM-FM sinusoidal modulations (e.g. vibrato-like sinusoids) the slope of ω_r will depend on the coupling between the modulating waveforms. Fig.1 shows $X = X(\omega\omega_r)$ and $\omega_r = \omega_r(\omega)$ for the peaks corresponding to three different sinusoidal cases: stationary, FM and AMFM. It is interesting to verify that although sinusoidal peaks expand in ω in presence of modulation, they have almost the same width with respect to the product $\omega\omega_r$. Therefore, the energy density concentration ($C_{\omega\omega r}$) will be determined completely by the peak shape along the $\omega\omega_r$ axis. As the peak shape against ω is already incorporated in ω_r through the gradient evaluation, we assume that:

$$C_{\omega\omega r} = g(\omega_r, \Delta\omega). \qquad (4)$$

If, for example, ω_r is fixed then $C_{\omega\omega r}$ will depend exclusively on the peak width $\Delta\omega$. Let us now apply this remark for characterizing noise peaks in the following subsection.

2.2. Noise

The peaks from the DTFT of the Gaussian noise must be characterized in a probabilistic sense. It is quite difficult to evaluate ω_r because of the random nature of the spectrum modulus. Nevertheless, we can estimate the mean peak width and hence evaluate (4) as a most probable value of the covariance. Let us recall that the DTFT modulus is formed by two independent identically-distributed Gaussian variables, namely, real and imaginary spectrum. Therefore, the DTFT modulus will have the Rayleigh distribution [7] with the variance dependent on the type of analysis window while the spectral phase will be uniformly distributed in $(-\pi, \pi)$. Next, we make use of the algorithm that estimates the number of local maxima in the envelope of a wide-sense stationary (WSS) band-pass zero mean Gaussian random process [8]. Briefly, it states that for a WSS band-pass zero mean Gaussian random process $I(t)$, written in polar form:

$$I(t) = R(t)\cos[2\pi f_0 t + \theta(t)] , \qquad (5)$$

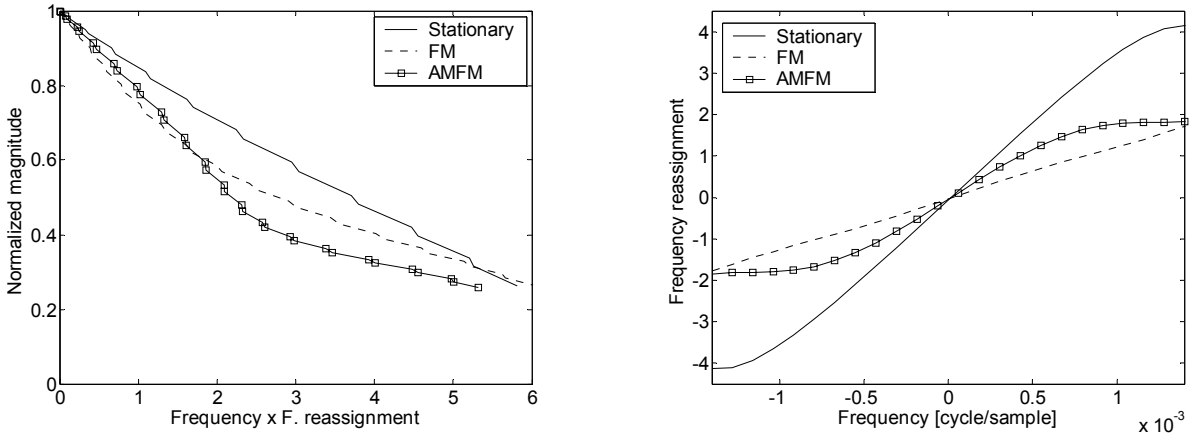

Figure 1- Representation of three different sinusoidal peaks versus $\omega\omega_r$ (left) together with the respective ω_r (right); the analysis window is Hanning.

every maximum of $R(t)$ corresponds to a zero with negative slope in $R'(t)$. Then, the expected number of zeros will be:

$$N = -\int_{-\infty}^{0} r'' P_{R'R''}(0, r'')\, dr'' \qquad (6)$$

where $R' = R'(t_o)$, $R'' = R''(t_o)$ and $P_{R'R''}(r', r'')$ is the joint PDF of those random variables. As the calculus for (6) is thoroughly explained in [8] we herein present only the final result:

$$N \approx \frac{1}{2\pi}\sqrt{\frac{\alpha_2}{\alpha_1}},$$
$$\alpha_1 = \mu_0\mu_2 - \mu_1^2$$
$$\alpha_2 = \mu_0\mu_4 + 3\mu_2^2 - 4\mu_1\mu_3$$
$$\mu_n = (2\pi)^n \int_0^{\infty} f^n W_I(f)\, df, \qquad (7)$$

where $W_I(f)$ is an arbitrary shape power spectrum of $I(t)$. If we now substitute in (5) $R(t)$ and $\theta(t)$ by the DTFT envelope and phase respectively, we obtain a band-pass stationary Gaussian process whose $W_I(f)$ will be completely determined by the type of analysis window (as the phase spectrum will always have the same distribution). Consequently, N will be expressed as a number of maxima per bin frequency. As (7) does not depend on the carrier frequency, we can put $f_o = 0$ and calculate N for various analysis windows (the corresponding data is given in Table.1). From Table.1 we can see that for a given analysis window, the mean peak width varies from 1.56 bins (rectangular) to 2.63 bins (Blackman), calculated as 1/0.64 and 1/0.43 respectively. In a view of that and recalling (4), we expect (independently of ω_r) the noise peaks to have in general considerably smaller $C_{\omega\omega r}$ compared to sinusoids. This indeed will be shown in the next section.

Table 1 - Estimated number of maxima per frequency bin for different analysis windows

Window	α_1	α_2	N [max/bin]
Rectangular	3.29	51.95	0.64
Bartlett	1.64	14.61	0.48
Hanning	1.28	9.00	0.43
Blackman	1.00	5.67	0.38

3. PEAK DISTRIBUTIONS

In order to apply the above concepts in musical signal analysis, we should first generate statistics for sinusoidal and noise peaks. It can be done by calculating the covariance for noise and sinusoidal peaks separately and building the corresponding distributions. In both cases we have first established a signal model, next we have calculated the windowed STFT and finally we have analyzed the corresponding peaks in the spectrum modulus and generated the corresponding distribution. For the noise distribution we have applied a Gaussian zero-mean amplitude-normalized signal and analyzed all the peaks in the STFT modulus. For the sinusoidal distribution we have proposed an AM-FM vibrato-like amplitude-normalized signal model:

$$x(n) = \cos\left[2\pi F_0 n + A_{FM}\sin(2\pi F_{FM} n + \pi/4)\right] \times \left[1 + A_{AM}\cos(2\pi F_{AM} n + \pi/2)\right] + r(n),$$

where $F_o = 0.2$, $F_{AM} = 1/(2L)$, $F_{FM} = F_{AM}/2$, $A_{AM} = 0.5$, $A_{FM} = 10$, L is the size of the analysis window and $r(n)$ is a Gaussian additive noise. The sinusoidal distribution was generated by collecting only the peaks corresponding to the main lobes in the STFT modulus. Let us make a few remarks about the sinusoidal signal model. The mixed-modulation sinusoid is probably the most critical case in detection of non-stationary sinusoids.

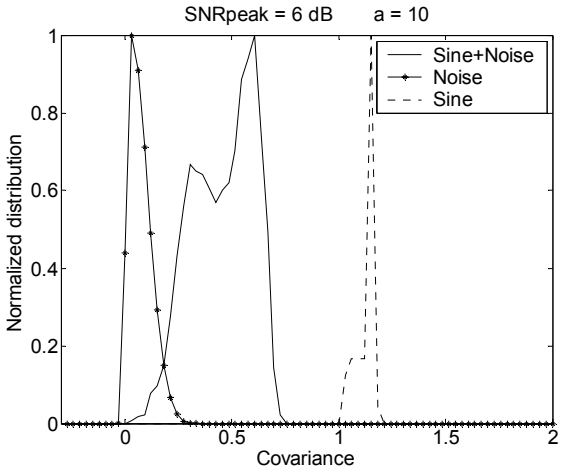

Figure 2 - Normalized distributions for three peak classes:
- sinusoidal AMFM with noise (SNR_{peak} = 6dB)
- sinusoidal AMFM without noise
- noise;
Analysis window is Hanning and a = 10.

Therefore, it is a good reflection of the reality in musical signals. Next, we observe that the size of analysis window is coupled to the period of the AM-modulation (and consequently to that of the FM). The reason for this lies in the assumed sinusoidal spectrum representation as a single peak whose properties vary in time. Otherwise the modulation sidebands would be resolved and thus the sinusoidal peak interpretation should be reformulated. Finally, the variance of $r(n)$ is given by the peak signal/noise ratio (SNR_{peak}) rather than through the global SNR. We decided on this kind of relation because it seems more appropriate for a narrow-band (peak) analysis. The distributions have been generated for the Hanning window and the parameter a = 10. From Fig.2. we see that in absence of $r(n)$ there is no overlap between the distributions. Nevertheless, even for the SNR_{peak} as low as 6dB the overlap is only approximately 5%. The decision threshold will usually be user dependent but more sophisticated way should be an automatic adjustment as a function of the SNR_{peak} and a.

4. EXPERIMENTAL RESULTS

We have analyzed a number of musical signals but due to the paper restrictions we present only one example, being a flute vibrato signal from the Utah University database. The signal was analyzed by a 2000-sample Hanning window and the spectrogram, the reassigned spectrogram and the covariance method are shown on Fig.3. The reassigned spectrogram has the best sinusoidal energy localization in the sinusoidal regions. However, it incorporates a lot of artifacts in purely noisy regions. The covariance method detects only the sinusoidal components with small errors due to the distribution overlap. In addition, it is almost free of spurious components in the noise regions.

5. CONCLUSIONS

We have shown that the reassignment process can be considerably improved by analyzing the time-frequency shift vector in a more general context. A set of spectral peak parameters like shape, width and frequency reassignment operator can be combined through the covariance expression for better sinusoidal energy localization. The proposed method is immune against spurious components in noise regions, it is quite robust to the additive noise and easy to implement. The main shortcoming lies in the by-hand adjustment of the decision threshold. Therefore, the future research will focus on the development of an algorithm for automatic threshold determination.

6. REFERENCES

[1] F. Auger, P. Flandrin, “Improving the readability of time-frequency and time-scale representations by the reassignment method”, IEEE Trans. On Signal Process. Vol.43, No. 5, pp. 1068-1089, May 1995

[2] E. Chassande-Mottin, P. Flandrin, F. Auger, “Supervised time-frequency reassignment”, IEEE Int. Symp. On Time-Frequency and Time-Scale Analysis, pp. 517-520, Paris 1996

[3] E. Chassande-Mottin, P. Flandrin, F. Auger, “On the statistics of spectrogram reassignment vectors”, Multimedia Systems and Signal Process. 9, 355-362, Kluwer Academic Publishers, 1998

[4] S. Hainsworth, M. Macleod, “On sinusoidal parameter estimation”, Proc. of the 6th Conference on Digital Audio Effects, London, UK, September 2003

[5] M. D. Macleod, “Fast nearly LM estimation of the parameters of real or complex single tones or resolved multiple tones”, IEEE Trans. Signal Processing, Vol.46, No.1, pp. 141-148, January 1998

[6] E. Chassande-Mottin, I. Daubechies, “Differential reassignment”, IEEE Signal Process. Letters, Vol.4, No.10, pp.293-294, October 1997

[7] R. Hippenstiel, “Detection theory: application and digital signal processing”, CRC Press, 2002

[8] A. Abdi, S. Nader-Esfahani, “Expected number of maxima in the envelope of a spherically invariant random process”, IEEE Trans. On Information Theory, Vol. 49, NO. 5, pp. 1369-1375, May 2003

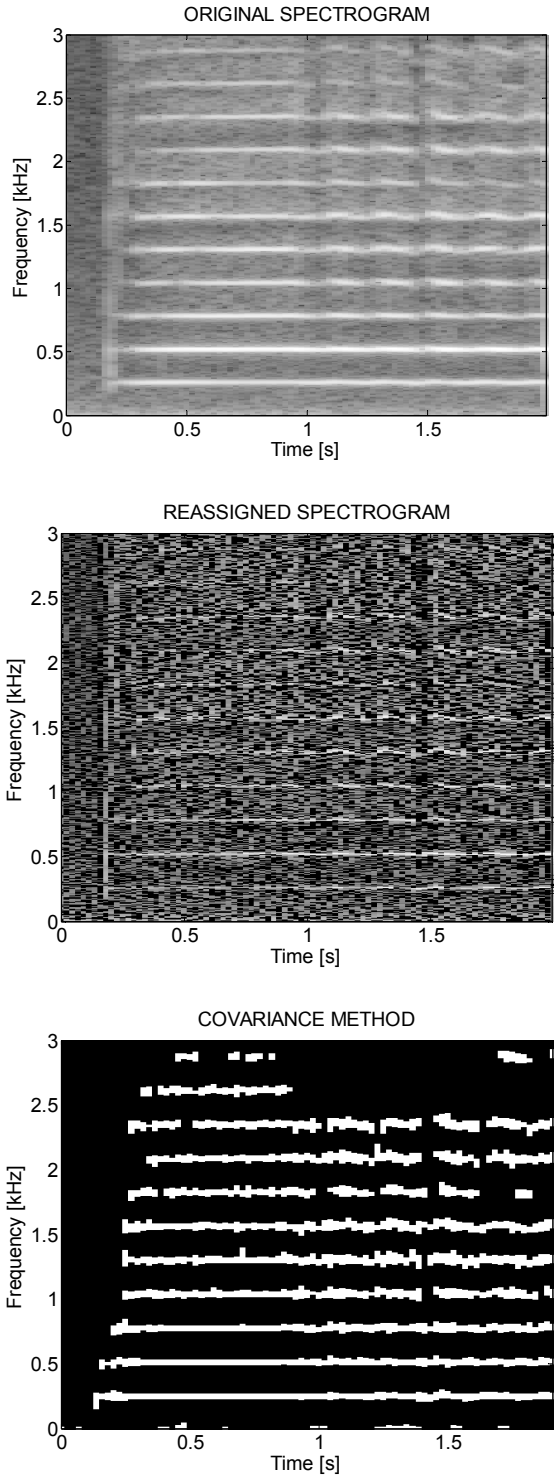

Figure 3 - Comparison among different time-frequency representations for a flute vibrato signal.